Temporal downscaling of irradiance data via Hidden Markov Models on Wavelet coefficients: Application to California Solar Initiative data

Jörg Wegener, Matthew Lave, Jennifer Luoma and Jan Kleissl

February 1, 2012

Department of Mechanical and Aerospace Engineering, UC San Diego
Project funded by the California Solar Initiative RD&D Program

Contents

1 Introduction and outline

The California Solar Initiative (CSI) rebate program requires a performance-based-incentive (PBI) payout for systems larger than 30 kW and makes it optional for smaller systems. This requires metering and monthly submission of 15 minute energy output to the payout administrator. We have obtained the 2010 quality-controlled CSI power output for 115 PV power plants in SDG&E, PG&E, and SCE territory. The data-set and quality control methods for system performance data are described in more detail in Itron and Kema (2010). The data were further quality controlled to remove effects of shading, low resolution power output, noise, inconsistent output (possibly due to disconnected strings), soiling, inverter clipping (undersized inverter), tracking, system downtime, and other effects not representative of irradiance (Luoma and Kleissl, 2012).

15 minute averaged data are not always sufficient. For example, to study power quality issues ('flicker') or impacts of PV sites on voltage regulation equipment, data at higher temporal resolution are required. At present, such data only exists at a few sites (e.g. less than five publicly available in California), but is mostly proprietary to power plant owners and system operators. Hence, one objective of the CSI RD&D program was to provide solar resource data at higher temporal resolution. We intend to mimic realistic behavior of solar irradiance observations at a higher temporal resolution than the recording interval of the data. Given merely the actual recordings, this so-called

'downscaling' produces realizations at a higher sampling rate that are expected to be *statistically* similar to actual recordings. In this CSI project, observations provided as 15 minute averages are used to generate a sample with one observation per 0.8789 seconds. This corresponds to a 'downscaling factor' of $2^{10} = 1024$.

In Section 2 of this document, the downscaling model is introduced and applied to clear-sky data from CSI network sites. Further, downscaling results are validated against Global Horizontal Irradiance (GHI) data at 1 second temporal resolution and converted to power plant output. A description of the provided data (Section 3) concludes this work.

2 Model overview

2.1 Model description

The key-properties of wavelet decompositions, namely simultaneous localization in time and frequency, are utilized to construct a frequency/scale dependent downscaling approach. A hidden Markov model in a tree arrangement is used to simulate a set of (partial) discrete wavelet coefficients, with the inverse transform leading to a downscaled version of the input signal. Due to the tree-like structure of the model, it is able to reflect certain inter-scale dependence structures (Crouse et al., 1998). Those structures can be associated with the impact of passing cloud shadows, when the impact permeates from longer (the cloud time scale) to shorter time scales. The wavelet coefficients at each time scale have to be calibrated using the variance at each time scale. A separate model is constructed to estimate variances at short time scales from the variance at 15 minutes. Based on observed variance decreases in 1 sec measured data, variance is reduced toward shorter time scales. The downscaling model is applied to clear-sky indexes obtained from CSI 15 minute averaged power data. In the downscaling process, the hidden Markov model is used to generate wavelet coefficients at the shorter scales. Results of this model are shown in Figure 1.

In the following, the method is validated through comparison of a coarser sampled and subsequently downscaled time series with the original (measured) series. The metrics are distribution functions of downscaled and original data and corresponding ramp rates. Moreover, estimated and sample variance versus time scale are compared. The data was recorded on the Uni-

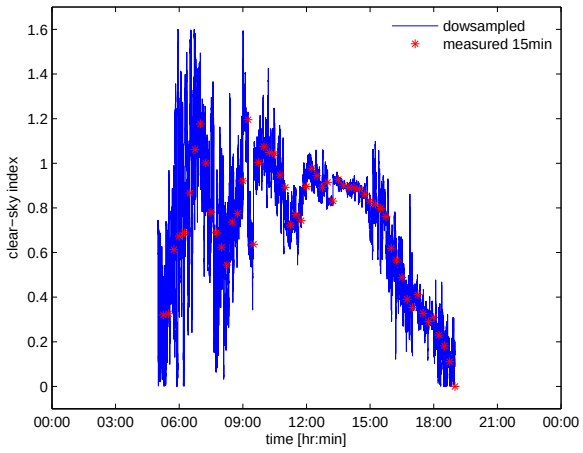

Figure 1: Downscaled clear-sky index at a sampling rate of one per 0.8789 seconds and CSI data at a rate of four per hour. Observations stem from a CSI site at 33.01° lat, -117.04° lon for June 1, 2010.

versity of California San Diego campus with a Licor 200Z silicon pyranometer and will be referred to as 'original','1sec measured' data or simply 'observations'. From the original data, we first generate a coarser sampled version analoguous to the 15 min CSI data. Somewhat loosely, these data are denoted as 'sampled at 15 minutes' temporal resolution. Through dividing the original time series into non-overlapping chunks of 1024 consecutive observations and taking the average of each chunk ('block-averaging'), we create a time series with a sampling interval of roughly 17 minutes. This coarsely sampled time series constitutes the point of departure for the downscaling procedure. Downscaling results are in turn compared to the original 1 sec data.

2.2 Application to clear-sky index data

Observations of solar irradiance usually show non-stationary behavior in the course of one day and little persistence on a day to day basis. These characteristics represent a challenge to every downscaling procedure. We illustrate our approach with two examples from the aforementioned data set with 1 second temporal resolution. Two days of data with broken and fluctuating cloud cover exemplify some typical non-stationary characteristics: Figure 2 (left)

shows observations from an overall clear day with scattered clouds appearing midday and intensifying throughout the evening. While the 15 minute average remains relatively constant during the day, the variance changes with time.

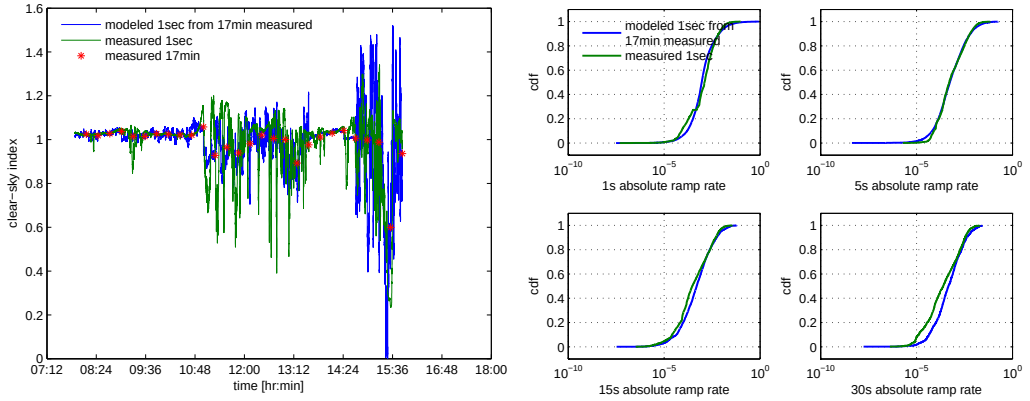

Figure 2: *Left:* Measured 1sec, block-averaged, and downscaled clear-sky index at the UC San Diego campus for January 5, 2009. *Right:* Cumulative distribution functions of absolute ramp rates of original and downscaled data; note the logarithmic scale on the abscissa.

Figure 2 (left) shows that the downscaled data follows the overall pattern as well as changes in variance. The latter is a result of the Markov model being organized in separate trees, such that every tree 'adopts' the local variance behavior. It should be noted however, that the number of trees can not be made arbitrary large (as would be advantageous for highly non-stationary data), as it depends on the number of samples available to downscale from. Naturally, the accuracy of the downscaling procedure increases with increasing sample size of observations.

The cumulative distribution functions (CDFs) of absolute ramp rates (Figure 2, right) are in very good agreement, mostly as a result of the good agreement between the variances as functions of scale (Figure 3, right). Finally, the estimated probability density functions (PDFs) of observed and downscaled data are shown to be in good agreement (Figure 3, left).

In Figure 4 (left) thick cloud cover in the morning transitions to clear sky after about 0900 h. During the cloudy period, both clear sky index and variance transition from low to high values as the cloud cover breaks up.

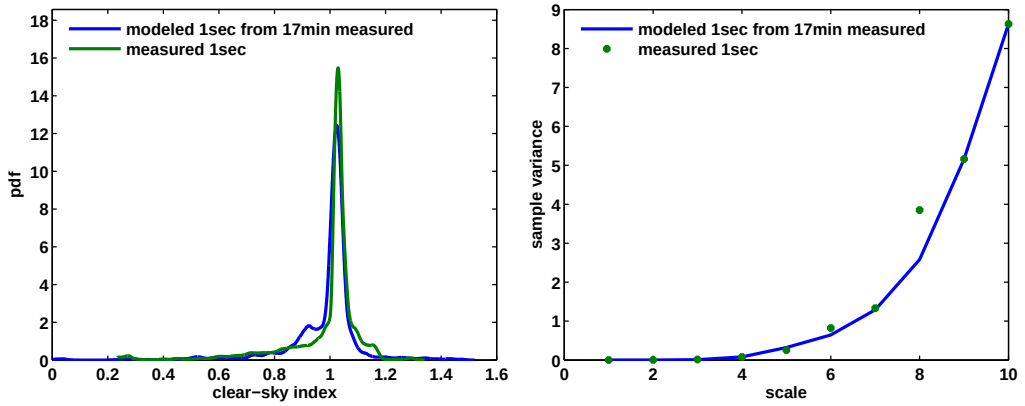

Figure 3: *Left:* Probability density functions for original and downscaled data shown in Figure 2. *Right:* Estimated (blue) and sample variance (green) per scale, where time = 2^{scale} sec. Note that the variances at shorter time scales are calibrated using the variances at scale number 9 and 10 (8.5 and 17 minutes); so, variances of wavelet coefficients in scale number 9 and 10 coincide with the original.

As can be noted from the graph, the downscaling model reproduces those regimes quite well.

The accuracy of ramp rate CDFs (Figure 4, right) is slightly worse compared to January 5, 2009, presumably due to a slightly worse fit of the variances per scale in Figure 5 (right). The PDFs in Figure 5 (left) show very good agreement between observations and downscaled data.

2.3 Conversion to power plant output

Besides clear sky indexes, normalized power output for typical CSI systems is also generated. The downscaled clear-sky index described in Section 2 exhibits the statistical behavior of single point measurements. However, since PV sites cover an area of a few to 10,000 square meters, the aggregate power output variability is reduced when compared to point measurements (Lave et al., 2011). To estimate the effect of geographic smoothing, we apply a wavelet-based variability model (WVM) as described in Lave and Kleissl (2012). Firstly, the downscaled (unit-less) clear-sky index is multiplied by power output estimated for clear skies to obtain power output in Watt.

The WVM estimates the amount of geographic smoothing at various

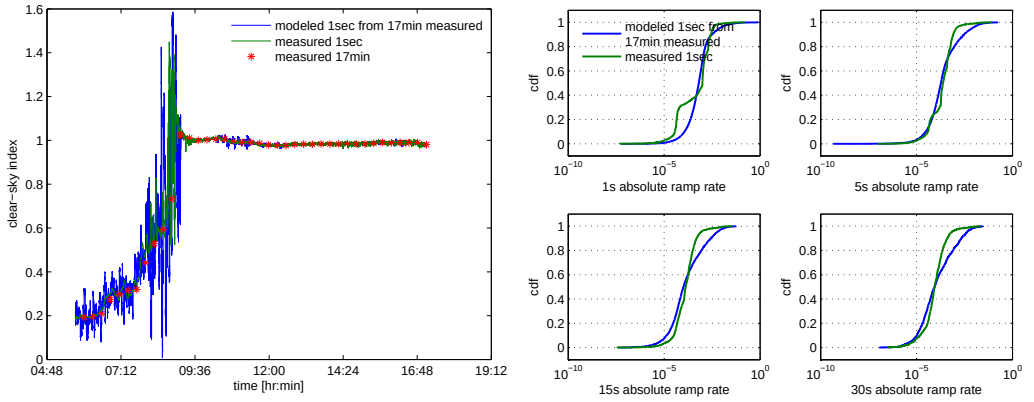

Figure 4: *Left:* Measured 1sec, block-averaged and downscaled clear-sky index at the UC San Diego campus for July 7, 2009. The latter is derived from block-averages of the original data, shown as red stars. *Right:* Cumulative distribution functions of absolute ramp rates of original and downscaled data with logarithmic scale on abscissa. Note that the 'bump' in the observed 1 sec ramp rate is an artifact of the precision of the sensor.

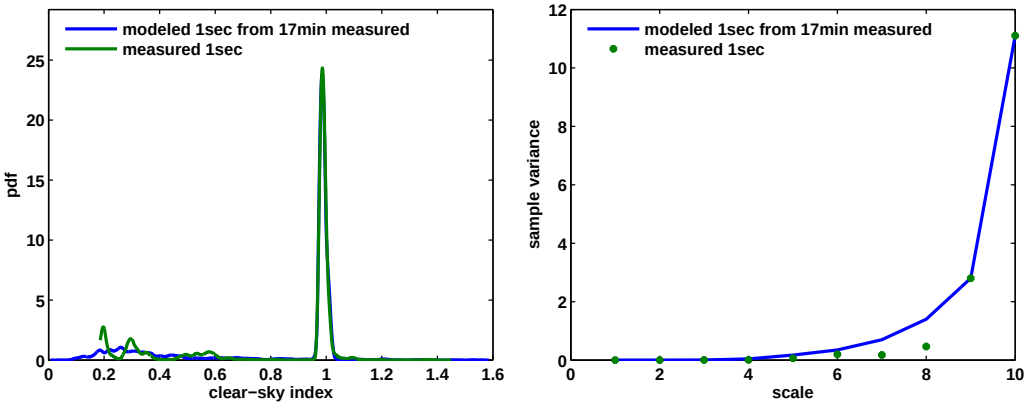

Figure 5: *Left:* Probability density functions for original and downscaled data shown in Figure 4. *Right:* Estimated (blue) and sample variance (green) per scale.

timescales by taking into account the dimensions of the system and their effect on correlation in power output between individual panels. A larger system will experience less correlation between the output of individual pan-

els, resulting in a smoothed output. Lave and Kleissl (2012) showed that the correlation as a function of distance and timescale was highly consistent across different sites and days if scaled by a coefficient A. Lower values of A indicate less correlation and hence more smoothing. Likely due to a difference in average cloud speeds, A was found to be larger for inland sites. For each day, we draw A from a uniform distribution on $[1, 3]$ for coastal and uniform on $[3, 5]$ for inland sites. Figure 6 shows sites categorized as coastal and inland, respectively; and Table 1 presents an overview of site location and PV plant capacity.

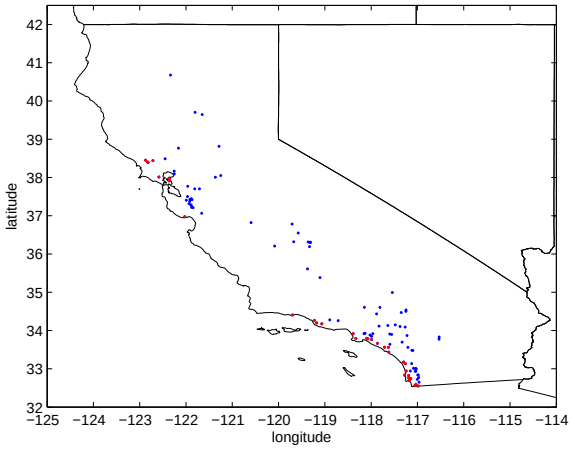

Figure 6: Geographical locations of CSI sites. Sites are classified as coastal (red) or inland (blue).

Applying the WVM to the downscaled power data does not affect the variability on longer time scales as shown in Figure 7 (left). However, a shorter snapshot (Figure 7, right) reveals some smoothing due to spatially upscaling the point observations to the power plant. In this example, the PV systems are relatively small such that smoothing occurs at short timescales (a few seconds), only. This effect is nonetheless critical to consider when computing e.g. short-time ramp rates of system power output as it reduces power quality impacts over short time scales.

Location			Location			Location		
Lat [°N]	Lon [°W]	Cap [kW]	Lat [°N]	Lon [°W]	Cap [kW]	Lat [°N]	Lon [°W]	Cap [kW]
36.8	119.7	940	36.2	120.1	976	33.5	117.1	7
37.3	121.9	346	38.0	122.4	133	34.2	119.1	984
37.2	121.9	333	38.4	122.8	37	34.4	117.9	7
38.0	121.4	339	34.4	119.7	193	33.8	116.5	5
35.4	119.1	323	36.3	119.3	77	34.5	117.2	6
37.3	121.9	640	33.9	118.0	66	34.3	118.7	85
38.0	122.6	358	33.9	117.6	321	34.2	119.2	52
37.5	122.0	145	33.6	117.2	292	35.0	117.5	37
37.2	121.9	197	33.6	117.7	344	34.5	117.3	20
38.8	121.3	1000	34.1	117.8	225	33.8	116.5	12
38.2	122.3	231	34.6	118.2	214	33.2	117.3	284
37.4	121.9	553	33.6	117.6	200	32.7	117.0	140
40.7	122.3	8	34.1	117.3	996	32.8	117.0	323
37.9	122.4	858	36.3	119.3	328	32.6	117.0	352
38.5	122.5	284	34.5	117.4	347	32.6	117.0	39
37.1	121.7	5	33.9	117.2	998	33.1	117.3	53
37.3	121.9	702	33.9	118.1	48	33.6	117.6	133
38.5	122.9	267	33.7	117.9	74	33.1	117.1	306
36.6	119.6	8	34.3	118.9	245	32.9	117.1	214
37.8	122.0	999	34.3	119.2	350	33.4	117.6	46
38.4	122.7	60	33.8	118.1	193	32.8	117.0	48
37.4	121.9	185	33.9	118.0	226	33.0	117.0	326
37.4	121.9	418	33.5	117.1	321	32.6	117.0	507
37.4	121.9	160	34.2	117.5	214	32.9	117.0	502
37.4	121.9	102	33.9	118.0	221	33.0	117.1	999
37.4	121.9	168	34.1	117.6	348	32.6	117.1	573
37.0	122.0	65	33.8	118.3	117	32.7	117.1	3
38.1	121.3	14	33.7	117.3	49	32.8	117.3	2
38.0	122.4	73	36.3	119.7	494	32.9	117.2	3
37.4	122.0	77	33.9	118.4	522	32.8	117.2	2
36.8	120.6	32	36.2	119.3	449	32.7	117.0	4
38.1	122.3	470	36.3	119.4	10	32.9	117.2	4
38.4	122.8	44	34.6	117.8	9	32.8	117.0	7
38.8	122.2	37	34.1	117.4	373	33.0	117.0	6
37.7	121.7	458	33.9	118.2	2	32.8	117.2	5
39.7	121.7	74	33.9	117.6	7	32.8	117.2	7
39.7	121.7	323	33.8	118.1	6	32.7	117.2	4
39.7	121.8	240	37.7	121.8	620	33.0	117.0	9
35.6	119.4	281						

Table 1: Overview of CSI site location and capacity. Locations are rounded to one tenth of a degree.

3 Description of database and file format

Data are stored one day per file, 365 files per site and year. Each file consists of five comma separated value (csv) columns of floating point numbers. Each value is rounded to the decimal place as shown in the below example, corresponding to format string '%3d%3d%8.5f%7.5f%8.3f' in C-like syntax.

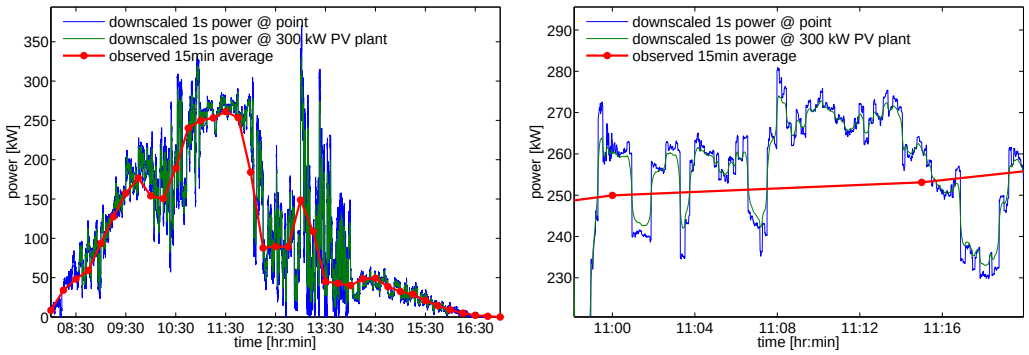

Figure 7: *Left:* Downscaled, downscaled with wavelet-based variability model (WVM), and 15min power output. *Right:* Detailed display of 20 minute outtake of left-hand figure. Observations stem from a 300 kW CSI site at 38.01° lat, -122.59° lon for January 3, 2010.

Lines are terminated by the line feed character (*A* in hexadecimal ASCII notation). The right-most column consists of PV plant power output data (cf. Section 2.3) given in units of kilowatt. Downscaled clear-sky index as described in Section 2.1 and displayed in Figure 1 is listed in the 4th column. The 3 left-most columns show the timestamp in hours (1st column), minutes (2nd column) and seconds (3rd column), respectively. Values in the third column are incremented by 0.8789 seconds. For example, the line

```
8, 30, 0.87890,0.09737,  27.969
```

shows a clear-sky index of 0.09737 and a power output of 27.969 kW at 8:30am and 0.8789 seconds.

File names are given in the following format

`lat_lon_capacity_year_day.csv`,

i.e. the concatenation of the respective site's geographical location, date of recording and site capacity in kW, delimited by underscores. The location is given as (signed) decimal degrees latitude and longitude, whereas the date is indicated by year and day of year (ordinal date). For example, the file

`033.5_-117.1_321_2010_001.csv`

contains one day (January 1, 2010) of data for the site at 33.5° latitude and -117.1° longitude with capacity 321 kW. All site locations are shown in Figure 6.

3.1 File size and handling of missing values

The number of lines per file varies as the amount of daylight hours changes with season. Hence, file sizes vary accordingly. An individual file covers the time period between sunrise and sunset. Missing values are recorded as 'NaN', Not-a-Number. In the rare case (less than 5 instances per year and site) of a *whole day of missing values*, a full 24 hour cycle of NaNs is covered. This constitutes the maximum file size of 3342336 bytes.

3.2 How to request data

Due to the large data volume, data are not posted online. Data is available upon request from http://solar.ucsd.edu/datasharing/ .

References

Crouse, M. S., Nowak, R. D., and Baraniuk, R. G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902.

Itron Inc., KEMA Inc. (2010). CPUC California Solar Initiative 2009 Impact Evaluation. *http://www.cpuc.ca.gov/NR/rdonlyres/70B3F447-ADF5-48D3-8DF0-5DCE0E9DD09E/0/2009_CSI_Impact_Report.pdf* .

Lave, M., Kleissl, J., and Stein, J. (2011). A wavelet-based variability model (WVM) for solar PV powerplants. Submitted to *IEEE Transactions on Sustainable Energy*, Special Issue on Solar Energy.

Lave, M., Kleissl, J. (2012). Testing a wavelet-based variability model (WVM) for solar PV powerplants. Presented at the IEEE Power & Energy Society General Meeting, 2012.

Luoma, J., Kleissl, J. (2012). CSI Data Quality Control.